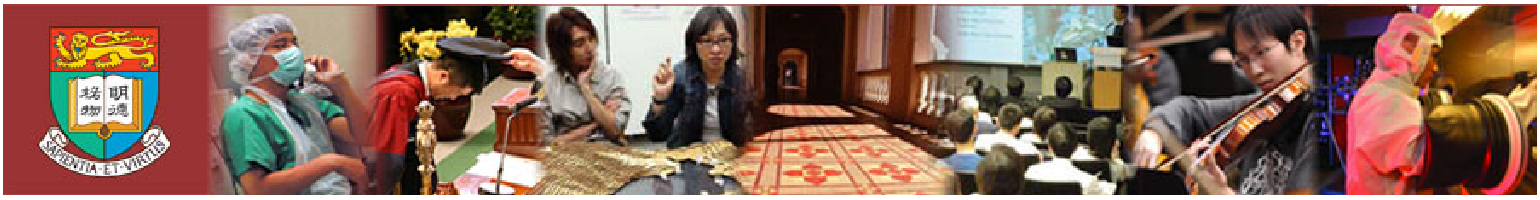

Title	Acoustical differences in vocal characteristics between Cantonese and English produced by Cantonese-English bilingual adult speakers
Author(s)	Chan, Yuk-kwan; 陳育君
Citation	
Issued Date	2010
URL	http://hdl.handle.net/10722/173698
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Acoustical Differences in vocal characteristics between Cantonese and English produced by Cantonese-English bilingual adult speakers

2006147783

A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2010.

Acoustical Differences in vocal characteristics between Cantonese and English produced by Cantonese-English bilingual adult speakers

Chan Yuk Kwan

A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2010.

Abstract

Vocal characteristics associated with 40 (20 male and 20 female) adult Cantonese-English bilingual speakers were examined. Fundamental frequency ($F0$) values and four parameters calculated from the long-term average spectrum (LTAS) including the first spectral peak (FSP), mean spectral energy (MSE), spectral tilt (ST) and high frequency energy (HFE) were obtained from connected speech samples produced in each language by the speakers. Acoustical measurements were measured with PRAAT and compared between languages and genders. The results indicated that female speakers had significantly higher $F0$ values in speaking English than Cantonese. Bilingual speakers exhibited comparable FSP values, but with significantly greater MSE value and lower ST value in Cantonese than English. The findings implied that, even with the same phonatory apparatus, language being spoken posed effect on speakers' voice quality.

Introduction

For any complex periodic signal, the fundamental frequency (*F*0) is the lowest frequency among all the harmonics. It is easily revealed after the signal is being analyzed using Fourier Transformation, with other frequencies being multiple integrals of it (Kent & Read, 2002). Physiologically, *F*0 reflects the rate of vocal fold oscillation during voicing, which is determined by the intricate activity of a number of laryngeal muscles and aerodynamic events. The rate of vocal fold vibration largely depends on the instantaneous configuration of the laryngeal structures, including the morphology and tension of the matrix of laryngeal muscles. People of different physiques are expected to demonstrate differences in the way their laryngeal systems are used and how their vocal folds vibrate.

Studies have investigated the acoustical characteristics of speech produced by individuals of different ethnicities in order to determine if difference in laryngeal function exists between different ethnic groups. This gives the implication to the application of normative speaking *F*0 data for clinical uses among different ethnic groups. These studies were set out to investigate if *F*0 characteristics are different among speakers of different ethnicities. However, discrepant results have been reported. Awan and Mueller (1996) compared the mean speaking *F*0 of White, African-American, and Hispanic kindergarteners. Results indicated that African-American children produced significantly lower mean speaking *F*0s than Hispanic children. It has been speculated that such *F*0 difference may be related to linguistic or anatomical differences among ethnic groups. Xue, Needley, Hagstrom and Hao (2001) carried out a study comparing speaking *F*0 obtained from a picture description task of African-American females and Euro-American females who were between 70 to 80 years old. It was observed that elderly African-American females had significantly lower mean *F*0 than Euro-American females. The authors stated that "caution must be taken when applying normative mean speaking *F*0 data collected solely from one ethnic group to clinical decisions of diagnosis and prognosis regarding the speaking *F*0s of other ethnic groups" (Xue et al., 2001, p.8).

However, Morris (1997) reported that there was no significant difference between African-American and White-American 10-year old boys in both reading and spontaneous speech data. It was concluded that the normative data of speaking *F*0 obtained from White-American boys can be generalized to African-American boys for clinical use. In addition, Sapienza (1997) found no significant difference between African-American and White speakers for *F*0 in vowel production. It was speculated that more similarities than differences existed in the vocal function between African-American and White speakers.

As speakers of different ethnicities generally spoke different languages in these studies, language difference might have contributed to the observed *F*0 differences. In order to find out whether those reported differences in speaking *F*0 was due to the interracial anatomical or linguistic difference, effect of language difference on acoustical characteristics has also been investigated. In a study comparing the English vowels produced by Mandarin, American-English and Hindi speakers, Andrianopoulos, Darrow, and Chen (2001) found that Mandarin-Chinese male and female speakers demonstrated significantly higher mean *F*0 values for isolated vowels than Caucasian and African-American speakers and Hindi Indian speakers. The authors attributed the differences to either the racial/ethnic difference or the language difference. As Mandarin is a tone language, it is expected to have greater range of *F*0 because of the varying nature of lexical tones. Altenberg and Ferrand (2006) argued that the use of isolated vowels in Andrianopoulos et al.'s (2001) study was invalid to represent typical continuous speech. They carried out an investigation of speaking *F*0 obtained from bilingual speakers producing continuous speech. Bilingual speakers of the same ethnicity were recruited in order to eliminate the interracial physiological difference. They compared the *F*0 values of English and Cantonese produced by nine bilingual English-Cantonese young women. Connected speech samples were collected and acoustically analyzed, and no significant difference in speaking *F*0 was found. Though continuous speech sample was obtained in Altenberg and Ferrand's (2006) investigation, only female subjects were included which rendered the results not generalizable. In addition, the bilingual speakers were believed to be more dominant in English with Cantonese was only used as a second language. Difference in result is expected should language dominance is reversed (i.e., Cantonese as the first language).

Based on the above discussion, speaking *F*0 can possibly be affected by: ethnicity/race of speakers and/or the language being used. How language alone influences speaking *F*0 is still unclear. The suggestion that tonal language is associated with higher speaking *F*0 than non-tonal language is not yet confirmed. Race/ethnicity may confound the study of how age and language influence speaking *F*0 due to the possible physical difference among speakers. In order to study how language alone affects speaking *F*0, the effect of race/ethnicity needs to be excluded. This can be done by studying speaking *F*0 associated with different languages produced by bilingual speakers who speak both languages (L1 and L2) with comparable proficiency. If speaking *F*0 differs between the two languages spoken by the same bilingual speakers, such *F*0 difference can only be attributed to language effect, but not physical difference associated with race/ethnicity or age.

Speaking *F*0 is related to vocal fold vibration, which is the sound “source” of speech production. Studying the source from the speech output is difficult as all output speech has been modified by the vocal tract filter, known as the transfer function (Kent & Read, 2002). The effect of vocal tract filter needs to be removed before one can study the vocal fold vibratory characteristics. To obtain more information of the source characteristics from the output speech, the technique of long term average spectra analysis (LTAS) of speech can be used. LTAS analysis can obtain spectral information from a relatively long speech sample, through which process the effect of vocal tract resonance can be eliminated and the vocal fold vibratory characteristics can be examined. LTAS averages all voiced sounds across a continuous speech sample, including only the sound generated from vocal fold and eliminating those which may adversely influence on the average spectrum, like voiceless consonant sounds and pauses. LTAS analysis has been widely used in research studies such as evaluating the techniques of voice therapy (Tanner, Roy, Ash & Buder, 2005), classifying voice qualities (Wendler, Pauhut & Kruger, 1986), examining sex and gender related differences on vocal characteristics (White, 2001), etc.

By averaging the short-term segmental fluctuations, LTAS analysis enables us to examine the vibratory characteristics of laryngeal source in an effective and reliable way. The effect of vocal tract resonance can be averaged out after completing LTAS spectral analysis which produces an approximated depict of the source in the frequency domain (Löfqvist & Mandersson, 1987). Therefore, it is an appropriate acoustic analysis tool to study the vibratory source characteristics of bilingual speakers in speaking Cantonese and English for comparison in the present study.

In summary, whether language alone affects the vocal characteristics is still unknown. The purpose of this study was to investigate the effect of language (Cantonese and English) on vocal characteristics in LTAS while eliminating the physical difference due to races by using bilingual speakers in the same ethnic group.

Method

Participants

Forty native Cantonese speakers (20 males and 20 females) were recruited. They were adult speakers with ages ranging from 19 to 24 years. All of them were Cantonese/English bilingual speakers. To ensure homogeneity, all participants met the following selection criteria: (a) the participants must be Cantonese-English bilingual speakers; (b) they were free of voice, speech and language problems, and they passed a hearing screening at 20 dBHL at 0.5, 1, 2 and 4 kHz bilaterally; and(c) they had no signs of upper respiratory infections and free from colds or other upper respiratory problems on the day of testing.

Procedures

All recordings were made in a sound-treated booth of the Speech Science Laboratory of the University of Hong Kong. All participants were allowed to ask questions about the study and told that they would be tape recorded. They were instructed to read aloud two passages, “North wind and the Sun” (see Appendix A) in Cantonese and “The Rainbow Passage” (See Appendix B) in English, in a comfortable manner. To avoid any order effect of language, the participants were randomly grouped into two with half of them began speaking in Cantonese, while the other half began with English. Prior to the actual recording, the participants were provided with a brief practice period in order to familiarize them with the recording environment and speech material.

During the recording, a microphone-to-mouth distance of 10cm was maintained. Speech samples were recorded via a high-quality microphone (SM58, Shure) and a preamplification system (MobilePre USB, M-Audio). The recorded speech samples were digitized at a rate of 44 kHz and stored in a computer for later *F*0 and LTAS analyses.

Data analysis

As the voiceless elements of each speech sample may corrupt data of voiced portions after averaging, silence periods and unvoiced segments in each speech sample were manually trimmed (Löfquist & Manderson, 1987). Each speech sample was submitted to cycle-by-cycle *F*0 and LTAS analysis using the PRAAT software program. After the *F*0 analysis, the numerical results were carefully examined and manually compared with the waveform. Outlying data due to artifact of the algorithm were visually inspected and excluded. Based on the *F*0 data, average speaking *F*0 values was calculated. LTAS spectra were quantified and four measurements were performed for comparison: (1) First Spectral Peak (FSP), (2) Mean Spectral Energy (MSE), (3) Spectral Tilt (ST) and (4) High Frequency Energy (HFE).

FSP (in Hz) is a measure of first amplitude peak across the LTAS display (Goberman & Robb, 1999). It is assumed to represent the average *F*0 of the speech sample. It follows that vocal fold stiffness may determine the FSP values (Fuller & Horii, 1988). MSE was the calculation of the average amplitude value across the frequency range of 0 - 8 kHz, and it was expressed as a negative value relative to the highest amplitude peak within the frequency range. It was shown to be correlated with the laryngeal tension (Fuller & Horii, 1998). ST is the ratio of the energy between 0 and 1 kHz (frequency region of *F*0 and lower harmonics) and the energy between 1 and 5 kHz (frequency region of upper harmonics). It is a representation of the rate at which amplitudes of the harmonics decline in the LTAS contour. Therefore, a low ST value means that the upper harmonics have more energy comparing to the lower harmonics. Mendoza, Munoz and Naranjo (1996) found that a low ST value in adult has been shown to correspond to hyperadduction vocal folds behavior and vice versa. HFE is the sum of amplitudes between 5 kHz and 8 kHz and was thought to be related to the presence of noise during phonation (Mendoza et al., 1996).

Two-way analyses of variance (ANOVAs) were conducted to compare the mean *F*0, First Spectral Peak (FSP), Mean Spectral Energy (MSE), Spectral Tilt (ST) and High Frequency Energy (HFE). Post-hoc multiple comparisons were used where necessary.

Reliability

To eliminate the possible human bias introduced during manual exclusion of the voiceless segments, intrajudge and interjudge reliability were measured. Accurate manual editing of the speech samples was critical and unavoidable for performing LTAS calculations. A set of 20% of the entire speech samples was randomly selected for re-editing, once by the primary investigator and once by a second investigator. Intrajudge reliability measure was obtained by comparing the duration values of the first and second edited speech samples obtained from the primary investigator. Interjudge reliability was obtained by comparing the values obtained from the first and second investigators. Results indicated that intrajudge relative percentage error was 0.26% and interjudge relative percentage error was 0.33%. Both measures indicate that the editing done by the primary investigator was consistent and reliable.

Results

Fundamental Frequency (*F*0)

The fundamental frequency (*F*0) values associated with English and Cantonese produced by Cantonese-English bilingual male and female speakers are shown in Table 1. A 2 x 2 (language x gender) two-way ANOVA was carried out to examine the differences in *F*0 across language and gender. As significant language by gender interaction effect was seen [$F(1, 38) = 4.22$, $p < .05$], separate paired-sample t-test was carried out to further analyze the changes in *F*0 across languages within each gender group. It showed that average *F*0 of Cantonese of female is significantly lower than that of English [$t(19) = -0.158$, $p < .05$]. No significant difference in average *F*0 in language of male ($p > .05$).

Table 1.

The average F0 of male and female bilingual adult speakers in Cantonese and English.

		*F*0 (Hz)	
Group		Cantonese	English
Male	*M*	121.35	121.53
	SD	13.07	12.96
Female	*M*	218.47	223.91
	SD	21.30	21.56

First Spectral Peak (FSP)

The Cantonese and English first spectral peak (FSP) data for the group of male and female bilingual speakers are listed in Table 2. A 2 x 2 (language x gender) two-way ANOVA was conducted to evaluate whether the FSP values differed significantly between Cantonese and English (language factor), as well as between male and female (gender factor). There was no significant language by gender interaction ($p > .05$) or significant main effect for language ($p > .05$). Significant main effect for gender was found, with female speakers exhibiting higher first spectral peak values compared to male speakers [$F(1, 38) = 31.25$, $p = .000$].

Table 2.

Mean and standard deviation values of FSP (in Hz) and MSE (in dB) obtained from Cantonese and English from the male and female bilingual speakers.

		FSP(Hz)		MSE(dB)	
Group		Cantonese	English	Cantonese	English
Male	*M*	141.90	163.90	-15.75	-16.47
	SD	32.30	83.38	1.58	1.89
Female	*M*	240.90	235.40	-16.44	-17.44
	SD	39.40	64.48	2.30	2.30

Mean Spectral Energy (MSE)

The Cantonese and English mean spectral energy (MSE) data for the male and female speakers are listed in Table 2. A 2 x 2 (language x gender) two-way ANOVA was performed to evaluate whether the overall mean spectral energy differ across language and gender. There was no significant language by gender interaction ($p > .05$) or main effect for gender ($p > .05$). There was a significant main effect for language, with Cantonese showing a higher mean spectral energy value compared to English [$F(1, 38) = 17.065$, $p = .000$].

Spectral Tilt (ST)

The Cantonese and English spectral tilt (ST) values exhibited by the male and female speakers are listed in Table 3. A 2 x 2 (language x gender) two-way ANOVA was performed to investigate the differences in spectral tilt across language and gender. There was no significant language by gender interaction ($p > .05$) or main effect for gender ($p > .05$). Significant main effect was found for language, with English has a higher spectral tilt than Cantonese [$F(1, 38) = 4.239$, $p < .05$].

Table 3.

Mean spectral tilt (ST), high frequency energy (HFE) and fundamental frequency (F0) values of Cantonese and English produced by male and female bilingual speakers.

		ST		HFE(dB)	
Group		Cantonese	English	Cantonese	English
Male	*M*	0.42	0.43	15969	16157
	SD	0.05	0.05	7758	6607
Female	*M*	0.41	0.43	23119	21283
	SD	0.05	0.06	10872	9946

High Frequency Energy (HFE)

The Cantonese and English high frequency energy (HFE) values associated with male and female speakers are listed in Table 3. A 2 x 2 (language x gender) two-way ANOVA was conducted to examine for differences in high frequency energy across language and gender. There was no significant language by gender interaction ($p > .05$). There was no main effect for language ($p > .05$) implying no significant differences in high frequency energy across gender groups. Significant main effect was found for gender factor, with female bilingual speakers had a higher mean HFE than male bilingual speakers [$F(1, 38) = 5.172$, $p < .05$].

Discussion

Fundamental frequency (*F*0)

The present study investigated the possible effect of language on the vocal characteristics. The results showed that gender has a different effect on the *F*0 for different languages. Male bilingual speakers showed no significant difference in *F*0 between Cantonese and English, but significant difference was found in female bilingual speakers. Female bilingual speakers demonstrated a higher mean *F*0 for English than Cantonese. *F*0 represents the rate of vibration of vocal folds during voicing. It follows that, female bilingual speakers were speaking English with a faster average vocal fold vibration than when speaking Cantonese. The result was inconsistent with Altenberg and Ferrand's (2006) study in which no significant difference was found in speaking *F*0 of female English-Cantonese bilingual speakers. The inconsistency can be explained in at least two ways. First, there was difference in language dominance in the two studies. Bilingual speakers in Altenberg and Ferrand's (2006) study were dominant in English with Cantonese as their second language, while the bilingual speakers in the present study were using Cantonese as the dominant language. This may indicate that competency in one language may affect the other one in speaking *F*0, at least in the case of female speakers. Second, the use of different speech tasks might account for the *F*0 differences. Conversational speech samples were used in Altenberg and Ferrand's (2006) study, while the bilingual speakers were asked to complete a reading task in the present study. The use of different speech materials may in some way determine the *F*0 values. This suggestion is in line with Guimaraes and Abberton (2004), in which *F*0 values were found to be associated with different speech tasks produced by Portuguese speakers. The female participants showed a significantly higher mean *F*0 in reading task than conversational task. It indicates that different speech tasks may have different effect on speakers' *F*0. Another study on the speaking *F*0 of White- and African- American boys also reported the lower *F*0 in spontaneous speech than oral reading (Morris, 1997). Thus, it is possible that the use of different speech tasks yielded different results between the two studies.

In addition, the present results also indicated significant differences in mean *F*0 between male and female bilingual adult speakers, with female speakers associated with a higher mean *F*0. This was consistent with previous studies that gender had a significant main effect on *F*0 (Guimaraes & Abberton, 2004). This is apparently due to the anatomical differences between male and female larynges, as *F*0 values reflect how rapidly the vocal folds are vibrating during speech production. The differences in size and mass of vocal folds of the male and female bilingual speakers should be responsible for the distinctive average *F*0. Male speakers' vocal folds are larger and higher in mass than female speakers. Thus, they normally vibrate more slowly than females' and therefore they should yield a lower *F*0. In addition, the *F*0 of the female speakers was more or less 1.7 times those of male speakers in both languages. This finding agrees with previous research results (Guimaraes & Abberton, 2004).

Based on the findings of Andrianopoulos, Darrow, and Chen (2001) that Mandarin speakers had a significantly higher mean *F*0s for isolated vowels than English speakers, Altenberg and Ferrand (2006) speculated that the English-Cantonese bilingual speakers would exhibit a higher *F*0 in speaking Cantonese than English because both Mandarin and Cantonese are tone language, and "with *F*0 variations at the syllable and word level convey different lexical meaning" (Altenberg & Ferrand, 2006, p.95). Yet, they failed to find significant differences. The findings of the present study appear to contradict with Andrianopoulos et al.'s (2001) one with female speakers demonstrated a higher mean *F*0 in English than Cantonese and no significant difference was found for male speakers. Therefore, the relationship between tone and fundamental frequency is still worth to be explored. Further research should investigate mean *F*0 with bilinguals speaking other tonal languages to examine the possible relationship between tone and fundamental frequency.

First spectral peak (FSP)

FSP value is the frequency of the first amplitude peak of the LTAS display that represents the highest energy in the low frequency region. It has been suggested that FSP tends to reflect the average *F*0 across the phonatory sample (Fuller & Horii, 1988). Thus, it may be another valid way of obtaining average *F*0 information from a speech sample. However, according to the present results, it might not be accurate as FSP could be affected by the vocal tract effect and energy distribution in the LTAS may be "contaminated" by energy from resonance in the other low frequency region. Therefore, it may not exactly represent $F0$. In the present study, no significant difference in FSP was shown between the two languages, which is different from the results obtained for $F0$. This suggested that $F0$ is more sensitive to slight differences between Cantonese and English. On the other hand, the average FSP associated with female bilingual speakers was significantly higher than that with male speakers, which agrees with the result of $F0$. It can be concluded that, although FSP might not be a faithful representation of $F0$, it is still related to $F0$.

Mean Spectral Energy (MSE)

The MSE value was shown to be correlated with tension of laryngeal musculature in infants (Fuller & Horii, 1998). Fuller and Horii (1988) suggested that the MSE value was greater (being less negative) when the laryngeal tension increased. Both male and female bilingual speakers showed an average MSE value that was less negative in Cantonese than that for English. This indicates that, laryngeal tension might be higher when speakers spoke Cantonese when compared to English. Cantonese may sound more tense than English. This could be explained by the difference in the resonance patterns in Cantonese and English. In an accent modification study, Kerr (2000) found that the quality of resonance differed when speakers spoke in Cantonese and English. Speakers resonated more in the back of the mouth when speaking

Cantonese and more in the front of the mouth when speaking English. She stated that apart from the lingual movements, a tenser vocal folds and pharyngeal muscles were thought to contribute to the use of posterior resonance. It follows that the higher laryngeal tension during Cantonese production by the bilingual speakers could be accounted for by the use of posterior resonance when speaking Cantonese.

Spectral Tilt (ST)

The bilingual speakers demonstrated a lower ST value in Cantonese than English. It means that more energy at higher frequency region relative to lower frequency region was observed during production of Cantonese. According to Löfqvist and Mandersson (1987), a low ST value with significantly high frequency energy in the source spectrum corresponded to a hyperadductional vocal fold behavior. However, the present results showed that there was no significant difference in HFE between Cantonese and English. Thus, the conclusion on bilingual speakers' vocal folds was comparatively hyperadducted in speaking Cantonese could not be made solely based on the low ST value.

Hillenbrand and Houde (1996) stated that the ST value is correlated with the perception of breathiness. A lower spectral tilt indicated a more breathy voice quality. Besides, Mendoza, Valencia, Mufioz and Trujillo (1996) also attributed the lower spectral tilt in female Spanish speakers to greater level of aspiration noise. Based on this, a lower ST value of Cantonese may imply a greater aspiration noise when the bilingual speakers spoke Cantonese in comparison with

English. The bilingual speakers' Cantonese might have exhibited a more "breathy" quality than English. As a matter of fact, it is not uncommon to find differences in voice quality across languages. In a study of voice quality variability in bilinguals using LTAS analysis, Bruyninckx, Harmegnies, Llisterri and Pocholive (1994) showed that the voice qualities of the same person were significantly different in the two languages produced by Catalan/Spanish bilinguals, which is agreed by the present study. The existence of the comparable results that Cantonese-English bilinguals showed different voice quality in speaking either language indicated that voice quality varied across languages, even with the same phonatory apparatus. Investigation of voice quality in various languages should be carried out before one can more thoroughly generalize such findings.

Though the present results showed no significant difference in ST value between males and females, the average ST value of females was lower than males when using Cantonese which was their vernacular tongue. This indicated that women's voice quality was more breathy than men's. To some extent, this agrees with the findings of the previous research. Mendoza et al. (1996) investigated the possible differences in voice quality in Spanish men and women using LTAS and they discovered that women's ST value was significantly lower than the men's. It implied that women had a more breathy voice quality. The insignificant differences in this study might be due to the sampling difference between this study and Mendoza et al.'s (1996) one. In Mendoza et al.'s (1996) study, the participants' age range was from 20 to 50 years, compared with 19 to 24 years in the present study. This may additionally reflect the fact that the degree of perceived breathiness of women's voice increase with age. On the contrary, men did not differ significantly in the degree of perceived breathiness in aging (Linville, 2002). Considering such, it can be concluded that the age factor contributes to the degree of perceived breathiness differently for males and females. The inclusion of older participants in Mendoza et al.'s (1996) research might account for the significant difference in their study.

Despite the fact that the results from Mendonza et al.'s (1996) agreed with those of Klatt and Klatt (1990) that American female speakers demonstrated a breathier voice than male speakers, they suggested that various other subject groups should be studied before generalizing the findings. Thus, the results from the present study provide important evidence that nationality of participants might yield dissimilar effects, with Chinese female and male speakers not different significantly in the degree of perceived breathiness. Therefore, it is still necessary to study more other subject groups to find out if nationality is a factor affecting the voice quality.

Limitations of the Study

Interpretation and generalization of the findings should be done with care as there are sampling and methodological limitations in this research. Participants were recruited from the University of Hong Kong. Therefore, there was a selection bias that the sample might not be representative of the population. Besides, Oral reading task was administered in the present study that it might not be representative enough for reflecting speakers' spontaneous speech vocal characteristics. Vocal intensity was not controlled in the study. Speakers were allowed to read the passage in their comfortable normal reading voice. However, variations in vocal intensity within and between speakers might affect the values of *F*0, with *F*0 increases when loudness increases (White, 1998). Therefore, it is suggested to collect conversational samples and control on the vocal intensity for future research.

Conclusion

Female Cantonese-English bilingual speakers demonstrated a higher *F*0 in speaking English than Cantonese, while male counterparts showed comparable *F*0 in the two languages. This implied that gender might pose different effect on *F*0 when speaking different languages. The average FSP values of Cantonese and English are comparable that it is not sensitive enough when compared to *F*0 in detecting the subtle differences between the two languages. The Cantonese produced by bilingual speakers might sound more tense than English as they exhibited a greater (less negative) MSE value in Cantonese. It is believed to be related to the posterior resonance used by the speakers when speaking Cantonese. Besides, the ST value of Cantonese was found to be significantly lower than that of English. It represented that there was a greater acoustic loss at the glottis (more breathy in voice quality) in production of Cantonese. Based on the findings, an important indication is that even with the same phonatory apparatus, the language being spoken in somehow determines the voice quality. It interacts with gender that in females, language also controls *F*0. This gives a clinical implication that the application of the

*F*0 norm in voice therapy should be careful with the language being spoken and the speech tasks being administered. It is because *F*0 value varies accordingly even the *F*0 norms are from the same race as the patient. Further investigation on exploring the vocal characteristics of bilinguals speaking other languages and in different age range is promising. The interaction between gender and language on speakers' *F*0 could also be investigated because it was shown in the present study that only females' *F*0 differed between two languages. Moreover, the effect of tone language on mean *F*0 is still worth to be studied as there are discrepant findings revealed in this study and previous researches. In addition, perceptual study on different languages produced by bilingual speakers is also merit studying as it is reflected from this study that voice quality is different in language being spoken by the same person. Investigating acoustical differences in speech production with the use of bilingual individuals allows one to eliminate interracial physiological differences in the speech apparatus and any differences found in results could be attributed to linguistic influences. To sum up, language being spoken determines female speakers' *F*0 and all speakers' degree of breathiness of their voice. This is an important finding that for speakers in the same race having the same phonatory apparatus, language poses effect on one's vocal characteristics.

Acknowledgement

I would like to express my sincere gratitude to my supervisor, Dr. Lawrence Ng for his guidance, valuable comments and support throughout the study. Special thanks are given to all the participants and my beloved friends for their kind assistance and emotional support.

References

Andrianopoulos, M. V., Darrow, K. N., & Chen, J. (2001). Multimodal standardization of voice among four multicultural populations: fundamental frequency and spectral characteristics. *Journal of Voice, 15*(2), 194-219.

Altenberg, E. P., & Ferrand, C. T. (2006). Fundamental frequency in monolingual English, bilingual English/Russian, and bilingual English/Cantonese young adult women. *Journal of Voice, 20*(1), 89-96.

Awan, S. N., & Mueller, P. B. (1996). Speaking Fundamental Frequency Characteristics of White, African American, and Hispanic Kindergartners. *Journal of Speech and Hearing Research, 39*, 573-577.

Bruyninckx, M., Harmegnies, B., Llisterri, J., & Pocholive, D. (1994). Language-induced voice quality variability in bilinguals. *Journal of Phonetics, 22*, 19-31.

Fuller, B., & Horii, Y. (1988). Spectral energy distributions in four types of infant vocalizations. *Journal of Communication disorders, 21*, 251-261.

Guimaraes, I. & Abberton, E. (2004). Fundamental Frequency in Speakers of Portuguese for Different Voice Samples. *Journal of voice, 19*(4), 592-606

Goberman, A. M., & Robb, M. P. (1999). Acoustic examination of preterm and full-term infant cries: the long-time average spectrum. *Journal of Speech, Language and Hearing Research, 42*, 850-861.

Hillenbrand, J. & Houde, R. A. (1996). Acoustic correlates of breathy vocal quality: dysphonic voices and continuous speech. *Journal of Speech and Hearing Research,* 39, 311–321.

Kerr, J. (2000). Articulatory setting and voice production: Issues in accent modification. *Prospect, 15*(2), 4-14.

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis and perception of voice quality variations among female and male talkers. *Journal of Acoustic Society of America, 87*, 820-857.

Kent, R. D., & Read, C. (2002). *The acoustic analysis of speech*. Albany, NY: Thomson Learning.

Linville, S. E. (2002). Source Characteristics of Aged Voice Assessed from Long-Term Average Spectra. *Journal of Voice, 16*(4), 472-479.

Löfqvist, A., & Mandersson, B. (1987). Long-time average spectrum of speech and voice signals. *Folia Phoniatrica et Logopaedica, 21*, 221-219.

Morris, R. J. (1997). Speaking fundamental frequency characteristics of 8-through 10-year-old White- and African-American Boys. *Journal of Communication Disorder, 30*, 101-116.

Mendoza, E., Munoz, J., & Naranjo, N. (1996). The long- time average spectrum as a measure of voice stability. *Folia Phoniatrica, 48,* 57-64.

Mendoza, E., Valencia, N., Munoz, J., & Trujillo, H. (1996). Differences in voice quality between men and women: use of the long-term average spectrum (LTAS). *Journal of Voice,10,* 59–66.

Sapienza, C. M. (1997). Aerodynamic and Acoustic Characteristics of the Adult African American Voice. *Journal of Voice, 11*(4), 410-416.

Tanner, K., Roy, N., Ash, A., & Buder, E. H. (2005). Spectral moments of the long-term average spectrum: sensitive indices of voice change after therapy. *Journal of Voice, 19,* 211-222.

White, P. (1998). A study of the effects of vocal intensity variation on children's voices using long-term average spectrum (LTAS) analysis. *Logopedics Phoniatrics Vocology, 23,* 111-120.

White, P. (2001). Long-term average spectrum (LTAS) analysis of sex and gender-related differences in children's voices. *Logopedics Phoniatrics Vocology, 26,* 97-101.

Wendler, J., Rauhut, A., Kruger, J. (1986). Classification of voice qualities. *Journal of Phonetics, 14,* 483-488.

Xue, S. A., Needley, R., Hagstrom, F., & Hao, J. (2001). Speaking *F*0 characteristics of elderly Euro-American and African-American speakers: building a clinical comparative platform. *Clinical Linguistics and Phonetics, 15*(3), 245-252.

Appendix A

Cantonese reading passage

北風和太陽 (North Wind and The Sun)

有一天，北風和太陽爭論說，到底誰的本領高。當他們爭論的時候，有一個人經過，他正穿著一件厚厚私黑色外衣。

因此他們便說，看看誰能脫去那人身上厚厚的外衣。

北風首先狠狠的吹。可是他越吹得狠，那個人就越把外衣拉緊。所以，北風就放棄了。

一會兒後，太陽出來了。那個人很快便將外衣脫下來。北風只好承認太陽較他厲害。

Appendix B

English reading passage

The Rainbow passage

When the sunlight strikes raindrops in the air, they act as a prism and form a rainbow. The rainbow is a division of white light into many beautiful colors. These take the shape of a long round arch, with its path high above, and its two ends apparently beyond the horizon.

There is, according to legend, a boiling pot of gold at one end. People look, but no one ever finds it. When a man looks for something beyond his reach, his friends say he is looking for the pot of gold at the end of the rainbow.

Throughout the centuries people have explained the rainbow in various ways. Some have accepted it as a miracle without physical explanation. To the Hebrews it was a token that there would be no more universal floods. The Greeks used to imagine that it was a sign from the gods to foretell war or heavy rain. The Norsemen considered the rainbow as a bridge over which the gods passed from earth to their home in the sky.

Others have tried to explain the phenomenon physically. Aristotle thought that the rainbow was caused by reflection of the sun's rays by the rain. Since then physicists have found that it is not reflection, but refraction by the raindrops which causes the rainbows.

Many complicated ideas about the rainbow have been formed. The difference in the rainbow depends considerably upon the size of the drops; the width of the colored band increases as the size of the drops increases. The actual primary rainbow observed is said to be the effect of a super-imposition of a number of bows. If the red of the second bow falls upon the green of the first, the result is to give a bow with an abnormally wide yellow band, since red and green light when mixed form yellow. This is a very common type of bow, one showing mainly red and yellow, with little or no green or blue.